

University of Groningen

Hsc70-4 Deforms Membranes to Promote Synaptic Protein Turnover by Endosomal Microautophagy

Uytterhoeven, Valerie; Lauwers, Elsa; Maes, Ine; Miskiewicz, Katarzyna; Melo, Manuel N.; Swerts, Jef; Kuenen, Sabine; Wittocx, Rafa??l; Corthout, Nikky; Marrink, Siewert-jan

Published in:
Neuron

DOI:
[10.1016/j.neuron.2015.10.012](https://doi.org/10.1016/j.neuron.2015.10.012)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Uytterhoeven, V., Lauwers, E., Maes, I., Miskiewicz, K., Melo, M. N., Swerts, J., Kuenen, S., Wittocx, R., Corthout, N., Marrink, S., Munck, S., & Verstreken, P. (2015). Hsc70-4 Deforms Membranes to Promote Synaptic Protein Turnover by Endosomal Microautophagy. *Neuron*, 88(4), 735-748.
<https://doi.org/10.1016/j.neuron.2015.10.012>

Copyright

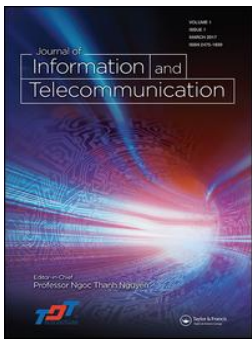
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



An analysis of rotation matrix and colour constancy data augmentation in classifying images of animals

Emmanuel Okafor, Lambert Schomaker & Marco A. Wiering

To cite this article: Emmanuel Okafor, Lambert Schomaker & Marco A. Wiering (2018): An analysis of rotation matrix and colour constancy data augmentation in classifying images of animals, Journal of Information and Telecommunication, DOI: [10.1080/24751839.2018.1479932](https://doi.org/10.1080/24751839.2018.1479932)

To link to this article: <https://doi.org/10.1080/24751839.2018.1479932>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 04 Jun 2018.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



An analysis of rotation matrix and colour constancy data augmentation in classifying images of animals

Emmanuel Okafor, Lambert Schomaker  and Marco A. Wiering

Institute of Artificial Intelligence and Cognitive Engineering (ALICE), University of Groningen, Groningen, The Netherlands

ABSTRACT

In this paper, we examine a novel data augmentation (DA) method that transforms an image into a new image containing multiple rotated copies of the original image. The DA method creates a grid of $n \times n$ cells, in which each cell contains a different randomly rotated image and introduces a natural background in the newly created image. We investigate the use of deep learning to assess the classification performance on the rotation matrix or original dataset with colour constancy versions of the datasets. For the colour constancy methods, we use two well-known retinex techniques: the multi-scale retinex and the multi-scale retinex with colour restoration for enhancing both original (ORIG) and rotation matrix (ROT) images. We perform experiments on three datasets containing images of animals, from which the first dataset is collected by us and contains aerial images of cows or non-cow backgrounds. To classify the Aerial UAV images, we use a convolutional neural network (CNN) architecture and compare two loss functions (hinge loss and cross-entropy loss). Additionally, we compare the CNN to classical feature-based techniques combined with a k -nearest neighbour classifier or a support vector machine. The best approach is then used to examine the colour constancy DA variants, ORIG and ROT-DA alone for three datasets (Aerial UAV, Bird-600 and Croatia fish). The results show that the rotation matrix data augmentation is very helpful for the Aerial UAV dataset. Furthermore, the colour constancy data augmentation is helpful for the Bird-600 dataset. Finally, the results show that the fine-tuned CNNs significantly outperform the CNNs trained from scratch on the Croatia fish and the Bird-600 datasets, and obtain very high accuracies on the Aerial UAV and Bird-600 datasets.

ARTICLE HISTORY

Received 30 November 2017



Accepted 20 May 2018

KEYWORDS

Image recognition; data augmentation; colour constancy; convolutional neural networks; feature descriptors

1. Introduction

Data augmentation (DA) has often been used in deep learning to increase the number of training images to obtain high classification accuracies. Previous approaches to data augmentation use cropping, rotation, illumination, scaling and colour casting for creating

CONTACT Emmanuel Okafor  e.okafor@rug.nl  Institute of Artificial Intelligence and Cognitive Engineering (ALICE), University of Groningen, Nijenborg 9, 9700 AK Groningen, The Netherlands

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

more training images. A recent research by Pawara, Okafor, Schomaker, and Wiering (2017) examined the classification performances of two convolutional neural network (CNN) methods (AlexNet and GoogleNet) with several DA techniques for different plant datasets. This research investigates the rotation matrix and colour constancy algorithms as methods for data augmentation with the objective to use one or more machine learning algorithms to classify images within three animal datasets.

Some researchers have considered rotating plant images in different angular positions while the effect of white or zero pixel values introduced during rotation of the images was not discussed (Ghazi, Yanikoglu, & Aptoula, 2017; Pawara et al., 2017), however, their research show that DA techniques can be used to reduce overfitting and improve the overall performance of the CNN models. A recent study investigated the relevance of the radial transform (Salehinejad, Valaee, Dowdell, & Barfett, 2018) as a method of data augmentation on character and medical multi-modal images. Additionally the research by Sladojevic, Arsenovic, Anderla, Culibrk, and Stefanovic (2016) attempts to develop a plant disease recognition CNN model with three image transformation techniques: affine, perspective and rotation.

In contrast to the rotation technique as mentioned earlier, the idea of colour constancy algorithms has widely been studied in image processing and computer vision as a method for enhancing the quality of an image while preserving the colour information of an object under varying illumination conditions. The authors in Rahman, Jobson, and Woodell (1996) and Jobson, Rahman, and Woodell (1997) have proposed a multi-scale retinex (MSR) method, which has the prowess to achieve excellent colour rendition and dynamic range compression as opposed to their previous works on the single scale retinex (SSR). An improvement was made in the MSR by the authors in Rahman, Jobson, and Woodell (2004), who incorporated colour restoration to produce a multi-scale retinex for colour restoration (MSRCR). Several improvements have been made on MSR to produce variants of the MSR algorithm. One of such methods is the combination of MSR with chromaticity preservation (Petro, Sbert, & Morel, 2014). Another modification on the MSR is the incorporation of the Autolevel algorithm that removes outliers, improves the contrast level within an image and shows computational improvements when used with a graphical processing unit (Jiang, Woodell, & Jobson, 2015).

However, the unification of colour constancy and rotation matrix algorithms as a method of data augmentation has received limited attention. This paper extends the research by Okafor, Smit, Schomaker, and Wiering (2017) by considering the proposed $n \times n$ rotation algorithm together with colour constancy techniques as methods of data augmentation. The proposed techniques are examined on two animal datasets (Croatia fish (Jaeger et al., 2015) and Bird-600 (Lazebnik, Schmid, & Ponce, 2005)) and an aerial image dataset collected using an unmanned aerial vehicle (UAV) (Okafor et al., 2017). The use of UAVs has a lot of potential for precision agriculture as well as for livestock monitoring. A previous study (Zhang & Kovacs, 2012) recommended that the combination between precision agriculture and remote sensing and UAV methods can be very beneficial for agricultural purposes. Other research (Katsigiannis, Misopolinos, Liakopoulos, Alexandridis, & Zalidis, 2016; López-Granados et al., 2016; Lukas et al., 2016) has examined this area with the use of UAVs for different tasks. A novel area of research is recognizing aerial imagery with the use of deep neural networks. The study in Lin, Cui, Belongie, and Hays (2015) demonstrates that the use of a CNN for ground-to-aerial localization yielded a good performance on some datasets.

Another interesting study is the use of deep reinforcement learning for active localization of cows (Caicedo & Lazebnik, 2015). Next to the task of localization, there exists some recent research on the use of UAVs for motion detection and tracking of objects. The study in Fang, Du, Abdoola, Djouani, and Richards (2016) analysed the merits of the use of optical flow with a coarse segmentation approach for aerial motion detection of animals from several videos. Furthermore, in Gonzalez et al. (2016) the authors extended the idea of using UAVs with object detection and tracking algorithms for monitoring wildlife animals. Another approach is detection and tracking of humans from UAV images using local feature extractors and support vector machines (SVMs) (Imamura, Okamoto, & Lee, 2016).

The idea of data augmentation has been successfully applied to UAV data as well. In Jeon et al. (2017), the authors studied augmentation of drone sounds using a publicly available dataset that contains several real-life environmental sounds. Furthermore, the research by Charalambous and Bharath (2016) explored the use of a DA method for training a deep learning algorithm for recognizing gaits. Another interesting use of data augmentation is the development of a model for 3D pose estimation using motion capture data (Rogez & Schmid, 2016). However, limited research has examined colour constancy as a method of data augmentation. The research by Galdran et al. (2017) proposed a DA method adapted for skin lesion analysis with neural networks with emphasis on the use of colour constancy to normalize the colour information of images within a training set. Moreover, a research has redeveloped colour constancy as a neural network regression technique for estimating the colour of a light source (Lou, Gevers, Hu, & Lucassen, 2015).

Most of the previous DA techniques transform a training image to multiple training images using techniques such as cropping, contrast, illumination, mirroring, colour casting, scaling and rotation. In this paper, we extend the DA method proposed in Okafor et al. (2017) that transforms a single input image to another image containing $n \times n$ rotated copies of the original (ORIG) image. This method enhances the amount of information in an image. Additionally, this paper investigates the use of two well-known colour constancy methods (MSR and MSRCR) for creating more samples of both original and rotation matrix versions of three datasets: Aerial UAV (Okafor et al., 2017), Croatia fish (Jaeger et al., 2015) and Bird-600 (Lazebnik et al., 2005). The objective of this paper is to use CNNs to assess the classification performance on several variants of the used datasets. Moreover, our study inspects if the novel DA methods lead to higher classification accuracies when combined with different machine learning techniques such as CNNs or classical feature descriptors on a novel dataset containing aerial images of animals.

Contributions: This paper describes a novel DA technique (Okafor et al., 2017) that transforms a train or test image into a novel single image with multiple randomly rotated copies of the input image. To combine the different rotated images, the proposed method puts them in a grid and adds realistic background pixels to glue them together. This approach presents some merits: (1) it provides more informative images which may aid to yield higher accuracies and (2) the method can also be used to perform data augmentation on test images in the operational stage. The utility of the proposed approach is evaluated by using a CNN which is derived from the original GoogleNet (Szegedy et al., 2015) architecture by keeping only several inception modules. For training this CNN, we evaluate if there are differences in using the cross-entropy loss function (softmax classifier) compared to using a hinge loss function. Furthermore, we compared the CNNs to several classical computer vision techniques using ORIG images and DA images. All techniques were

used to investigate the recognition accuracies of aerial images of cows in natural scenes, for which we created our own dataset with a UAV.

Additionally, this paper investigates the use of well-known colour constancy techniques (MSR and MSRCR) for creating new image samples of both ORIG and the new rotation matrix (ROT) images on three datasets: UAV aerial images, Croatia fish (Jaeger et al., 2015) and Bird-600 (Lazebnik et al., 2005), with the aim to increase the amount of training image samples. This approach enhances the colour information of the images which could be very useful to get higher classification accuracies with the CNN. We train the CNN with the cross-entropy loss function and compare the classification performances of the colour constancy data augmentation (with ORIG/ROT), ORIG alone and ROT-DA alone on three datasets. The study also considers two broad forms of data augmentation based on their increase (colour constancy data augmentation) or no increase (ROT-DA alone) in the amount of training images.

The results show that the fine-tuned CNN with an appropriate selection of the grid resolution and angular bounds for the rotation algorithm combined with colour constancy methods yields the highest classification accuracies on most of the used datasets. Moreover, the results show that using fine-tuned CNN models with the proposed data augmentation (ROT-DA) technique on the Aerial UAV images leads to significantly better results than all other approaches. Finally, the results of our proposed approaches to data augmentation combined with the fine-tuned CNN significantly surpass previous results on the Bird-600 dataset (Lazebnik et al., 2005).

Paper outline: Section 2 describes the used datasets and the proposed DA techniques. Section 3 discusses the methods used for classifying the Aerial UAV dataset and two other animal datasets. Section 4 describes the CNN experimental setups and the results obtained from the various classification methods on the used datasets. Finally, the conclusion is presented in Section 5.

2. Datasets and data augmentation

This section entails the description of three datasets and describes two kinds of data augmentation which are evaluated in Section 4.

2.1. Datasets

2.1.1. Aerial UAV dataset

- (1) Dataset collection: We employed the DJI Phantom 3 Advanced UAV for collecting video frames of cows and natural backgrounds at different positions and orientations (Okafor et al., 2017). An illustration of the UAV is shown in Figure 1.

We applied manual cut-outs with a fixed size of 100×100 pixels to obtain positive samples of images that contain a cow, while we employed an automatic extraction of negative samples which have no presence of cows in the image. We flew the drone three times over different fields containing cows in order to obtain different samples. A summary of the three subsets of the obtained images with the amount of positive and negative samples, the video streaming time and the amount of unique objects is reported in Table 1. The unique objects denote cows that are



Figure 1. A photo of the UAV used for this study.

recorded at different time frames and therefore have different appearances in time. [Figure 2](#) shows some samples of images of our Aerial UAV dataset.

(2) Cross-set splits: We used cross-set splits whereby each recorded subset is considered as a separate fold. One subset is used for testing and the other subsets are used for the training set. This process is repeated for the three available subsets. The classical feature descriptors combined with supervised learning algorithms and the derived CNN technique are employed for determining the existence of cows in the natural images. We maintain the same dataset splits for all the experiments using the CNN and the feature extraction techniques.

2.1.2. Croatian fish dataset

This dataset was originally presented in Jaeger et al. (2015). It contains a total of 794 images and has 12 classes with a non-uniform distribution of the images per class. The authors reported an accuracy of 66.78% in their study using a CNN combined with a linear SVM classifier. We adopted a different split in our experiment because of the imbalance of the image samples within the various classes. We ensured that approximately half of the image samples were kept aside as test sets. [Figure 3](#) shows sample images of this dataset for each of the classes.

2.1.3. Bird-600 dataset

This dataset was originally presented in Lazebnik et al. (2005). The dataset contains a total of 600 images and has 6 classes with 100 individual image samples per class. We adopted a similar dataset distribution by keeping 50% of the total image samples as test set as reported in Lazebnik et al. (2005) in our experiments. The authors reported an accuracy of 92.33% in their study by using a probabilistic part-based method for texture and object recognition. [Figure 4](#) shows sample images of this dataset for each of the classes.

Table 1. Statistics of video records and annotated data for the aerial UAV dataset.

	Video ID	Time (s)	Unique objects	Positive samples	Negative samples
1	Subset 1	11	10	37	225
2	Subset 2	43	82	475	2094
3	Subset 3	22	10	50	1100

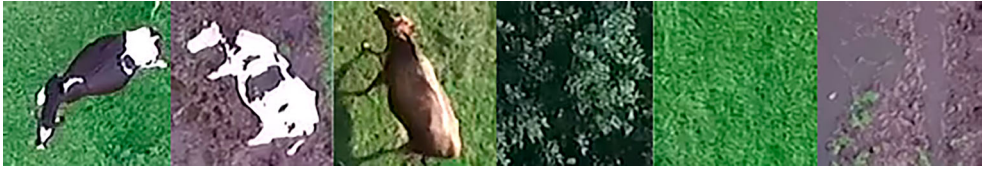


Figure 2. Sample images of the Aerial UAV dataset, showing the presence of cow (positive samples) and non-cow (negative samples). Please note that non-cow images are also quite diverse.



Figure 3. Sample images of the Croatia fish dataset showing each of the fish species (each column): *Chromis chromis*, *Coris julis* (female), *Coris julis* (male), *Diplodus annularis*, *Diplodus vulgaris*, *Oblada melanura*, *Sarpa salpa*, *Serranus scriba*, *Spicara maena*, *Spondylisoma cantharus*, *Symphodus melanocercus* and *Symphodus tinca* (Jaeger et al., 2015).

2.2. DA techniques

2.2.1. Multi-orientation data augmentation

We propose a new offline DA algorithm called ROT-DA that transforms an input image to a new single image containing multiple randomly rotated versions put in $n \times n$ cells. The use of a larger value for n leads to a new image containing more different poses. For the Aerial UAV dataset, the value of n was set to 4 in the experiments, because using higher values of n resulted in making the cow images look very small. On the other two animal datasets, we set $n = \{2, 4\}$ for Croatia fish while for the Bird-600, we set the value $n = \{1, 2\}$. An illustration of the proposed DA method and the overall classification system using the CNN is shown in Figure 5. The pseudo-code in Algorithm ?? explains the various transformations of the ORIG image to obtain the multi-orientation image. After inserting the images in the newly created image, background pixels are added to glue them together. This is done by using the nearest neighbour pixels around the edges of the images. We will also perform experiments with ROT-DA without rotations (ROT-DA-NR), but we do this only for the classical feature-based techniques.

Algorithm 1 Multi-Orientation Data Augmentation Algorithm

Input: Given images $I_i(x, y)$ from an input directory, where x, y denote the pixel row and column, and a grid size of $n \times n$.

Output: The data-augmented versions of the images.

- 1: **procedure** CONSTRUCT A FILELIST WITH N IMAGES FROM AN INPUT DIRECTORY.
- 2: **for** each image $I_i, i \in N$ **do**
- 3: Initialize the total number of cells $n \times n = M$
- 4: **for** each image I_i , for all cells $m \in M$ **do**
- 5: Define the size of the image resolution.
- 6: Compute a pad-size $I_q = \text{ceil}((\text{size}(I_i))/2)$.
- 7: Compute a pad-array I_p using a pixel replication padding technique, given I_i, I_q , pad value set to 'replicate' and the pad direction set to 'both'.
- 8: Rotate I_p with a random angle within the bound $[1^\circ, 180^\circ]$, this yields a new image I_r .
- 9: Adjust the image I_r to I_a such that undesired background introduced during rotation is filled with artificial pixels from the nearest neighbour pixels.
- 10: Concatenate each I_a into M cells.

11: $I_c = [I_a(k) \dots I_a(k+n-1); \dots; \dots I_a(M=n^2)]_{n \times n}$ Given that $k=1, \forall M$ cells, the ellipses (...) denote the column cells entries containing rotated sub-images, and the semicolon (;) in this study represents the start of a new row. Note that each cell in the $n \times n$ grid of cells contains a rotated copy of the input image $I_a(k)$ in a reduced size.

12: **end for**

13: Convert the cell structure of I_c into a matrix I_m .

14: Resize the image I_m to 250x250 pixels.

15: Store each $I_m(i)$ into an output directory

16: **end for**

17: **end procedure**



Figure 4. Sample images of the Bird-600 dataset for each of the bird species (each column): egret, mandarin, owl, puffin, toucan and wood duck (Lazebnik et al., 2005).

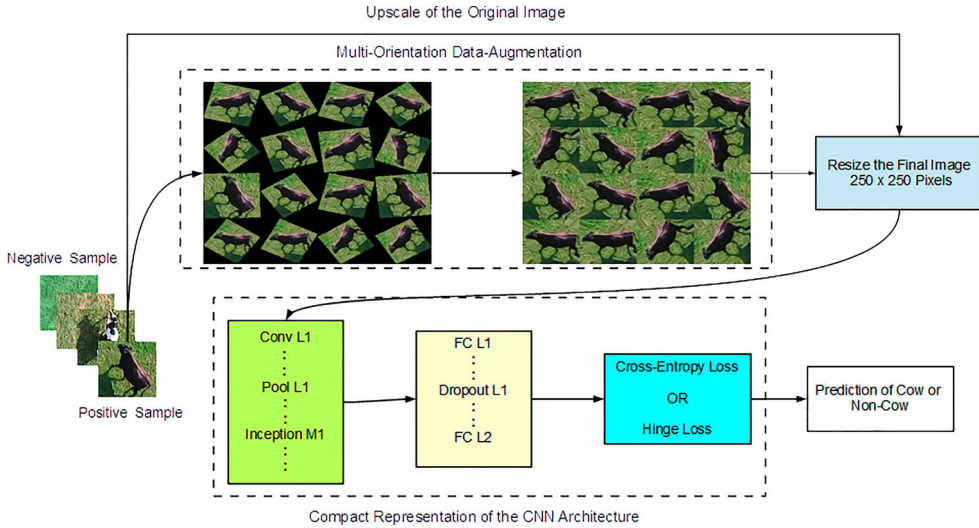


Figure 5. Block diagram illustrating the proposed method and overall system using the CNN. The column (':') symbol between different layers represents the connections of neural network layers within the derived CNN architecture. The data-augmented image on the top left is a multi-orientation image without padding and the image on the top right is the resulting multi-orientation image with padding.

2.2.2. Colour constancy data augmentation

Colour constancy is the perception of an object which ensures that perceived colours of objects remain relatively constant under various variations in illumination conditions. This area of study has found relevance in image processing and computer vision. Colour constancy uses contrast/lightness enhancement and colour rendition for improving the quality of an image. Most colour constancy algorithms use the retinex theory. The idea of the retinex theory was proposed initially by Land and McCann (1971). The research in Provenzi, Marini, De Carli, and Rizzi (2005) provided the basis for understanding the retinex algorithm from a mathematical standpoint. Our study examines two kinds of MSR algorithms.

- (1) Multi-scale retinex: This algorithm was proposed by Rahman et al. (1996) and Rahman et al. (2004). The algorithm provides a trade-off between colour rendition and local dynamic range (Petro et al., 2014). MSR computes the weighted sum of the outputs from various SSR. According to Jobson et al. (1997), an MSR image can be computed as

$$f_{msr_k}(x, y) = \sum_{m=1}^M W_m f_{m_k}(x, y) \quad (1)$$

$$f_{m_k}(x, y) = \log(l_k(x, y)) - \log\left(\sum_{(x, y)} C_m \exp\left[\frac{-(x^2 + y^2)}{2\sigma_m^2}\right] l_k(x, y)\right), \quad (2)$$

where f_{m_k} is the SSR output for \mathbf{M} scales, W_m denote the weights for each scale variable, $W_m = 1/3$, the maximum number of scales is $M = 3$ because the number of the RGB image channels is equal to the number of scales, C_m represents the normalization factor and $l_k(x, y)$ denotes the image pixel coordinates for a given colour band k . The $\sigma_m \in \{15, 80, 250\}$ are the standard deviations of the Gaussians for each of the scales. We adopted the same parameters as used in Jobson et al. (1997) and Petro et al. (2014), because they also perform well in our study. Furthermore, we further computed the $f_{msr_k}(x, y)$ by using the mathematical expression proposed in Moore, Allman, and Goodman (1991), where each colour channel is modified by the absolute minimum and maximum of the RGB colour channels. This can be computed as

$$f_{msr_k}(x, y) = 255 \frac{f_{msr_k}(x, y) - \min_k(\min_{(x, y)} f_{msr_k}(x, y))}{\max_k(\max_{(x, y)} f_{msr_k}(x, y)) - \min_k(\min_{(x, y)} f_{msr_k}(x, y))}. \quad (3)$$

- (2) Multi-scale retinex with colour restoration: Jobson et al. (1997) and Rahman et al. (2004) initially proposed the MSRCR algorithm. An MSRCR image f_{msrcr_k} can be computed by the product of colour restoration functions C_k of the chromaticity and the MSR outputs. The modified version of the MSRCR $f_{msrcr_k}(x, y)$ from the research in Petro et al. (2014) can be computed as

$$f_{msrcr_k}(x, y) = \lambda(C_k(x, y)f_{msr_k}(x, y) + \beta) \quad (4)$$

$$C_k(x, y) = \log\left[\alpha\left(\frac{l_k(x, y)}{\sum_{k=1}^K l_k(x, y)}\right)\right], \quad (5)$$

where α controls the strength of the non-linearity and λ is a constant. For the MSRCR experiment, α is set to 125 while λ is set to 0.8 and \mathbf{K} represent the total number of spectral bands ($\mathbf{K} = 3$) while β is set to 46.

Proposed colour constancy data augmentation: This study examines the possibility of using the ORIG or ROT images that are fed as input to the MSR or MSRCR algorithm. This process can also be done vice versa by creating the colour constancy images and then pass them as inputs to the rotation matrix algorithm. The new images are then combined with either ORIG or ROT images to obtain either double or three times the effective size of the initial train-validation image dataset. Please note by three times, we mean combining ORIG+MSRCR-ORIG+MSR-ORIG or ROT+MSRCR-ROT+MSR-ROT. We carried out experiments using two animal datasets and the UAV dataset. Some samples of both the ORIG and ROT images with and without colour constancy are shown in [Figures 6, 7](#) and [8](#) for the Aerial UAV dataset, Croatia fish dataset and Bird-600 dataset respectively.

We carried out some considerations to the rotational bounds for the ROT-DA alone or colour constancy data augmentation with ROT images on the three datasets.

- (a) For the Aerial UAV and Croatia fish datasets irrespective of the order of the grid cells, we used a rotational angle in the range $[1^\circ, 180^\circ]$.
- (b) For the Bird-600 experiments, we considered two rotational conditions for 2×2 -ROT-DA which we defined in two versions:
 - (i) Version 1 (V1): The rotational angles for different image poses lie in the bound $[1^\circ, 180^\circ]$. This computation was carried out on 2×2 -ROT-DA alone and colour constancy data augmentation with 2×2 -ROT images separately.
 - (ii) Version 2 (V2): The rotational angles for different image poses lie in the bound $[-15^\circ, 15^\circ]$ and we exempted angle 0° in our computation; this is because we do not want to have the existence of the ORIG image twice in the new DA variants. This computation was carried out only on the colour constancy data augmentation with 2×2 -ROT images.

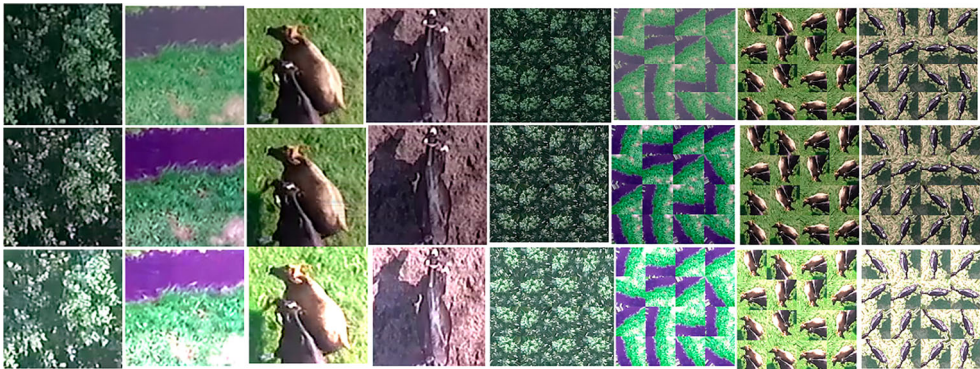


Figure 6. Examples of the ORIG and ROT-DA images from the Aerial UAV dataset. The first row accounts for the ORIG images (columns 1–4) and ROT-DA images (columns 5–8) without colour constancy. The second and third rows are the MSR and MSRCR versions for both the ORIG and ROT-DA images respectively. Our proposed rotation matrix algorithm eliminates zero pixel values generated due to rotation by filling it with nearest neighbour pixels. The colour constancy algorithm shows enhancement in the illumination and light intensities for each of the image samples.

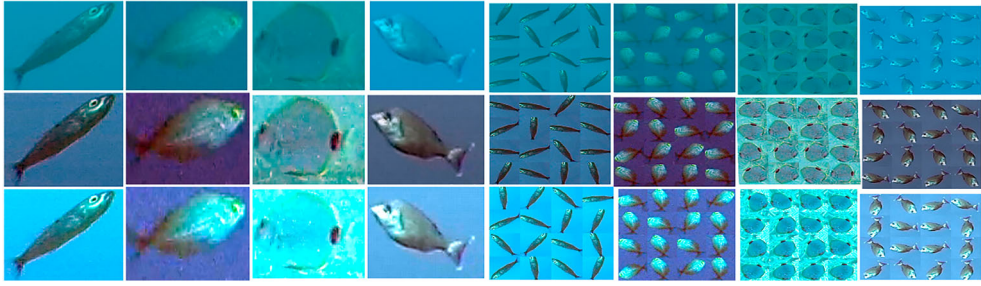


Figure 7. Examples of the ORIG and ROT-DA images from the Croatian fish dataset (Jaeger et al., 2015). The first row accounts for the ORIG images (columns 1–4) and ROT-DA images (columns 5–8) without colour constancy. The second and third rows are the MSR and MSRCR versions for both the ORIG and ROT-DA images respectively. The colour constancy algorithms also show improvement in the image resolution compared to the ORIG image samples.

- (c) On the Bird-600 experiments, we also considered the colour constancy data augmentation with 1×1 -ROT which used the same angular rotation bounds as in V2. This setup can be seen as a combined DA method of rotation and colour constancy.

3. Image recognition methods

3.1. Three inception module CNN architecture

This architecture is directly derived from the famous GoogleNet architecture as proposed in Szegedy et al. (2015). We eliminated all the layers after the inception 4a module, except for layers which lead to the first classifier and this is because the used datasets contain few classes (2, 6 and 12) for the Aerial UAV, Bird-600 and Croatia fish datasets respectively. Hence, we want to know how the reduced architecture can handle these problems. We will compare the reduced CNN architecture to the original GoogleNet on the Aerial UAV dataset. Another modification made with respect to the original GoogleNet architecture

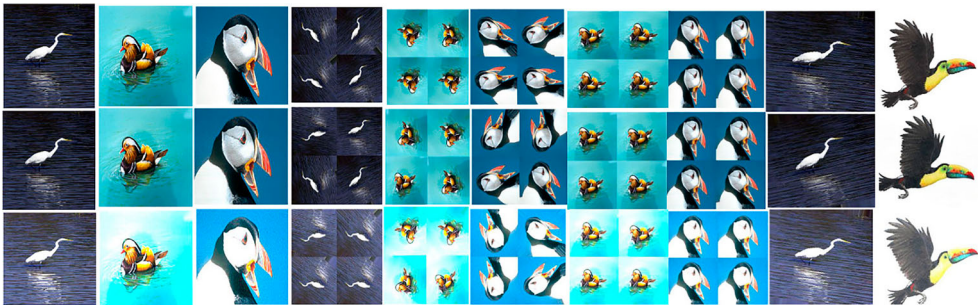


Figure 8. Examples of the ORIG and ROT-DA images. The first row accounts for the ORIG images (columns 1–3), 2×2 ROT-DA images using V1 rotation condition (columns 4–6), 2×2 ROT-DA images using V2 rotation condition (columns 7–8) and 1×1 ROT-DA images using V2 rotation condition (columns 9–10) all mentioned without colour constancy. The second and third rows are the MSR and MSRCR versions for both the ORIG and ROT-DA images respectively.

is the use of Nesterov's Accelerated Gradient Descent (NAGD) rather than using the conventional stochastic gradient descent (SGD) to update the weights in the deep neural network. The NAGD optimization update rule (Sutskever, Martens, Dahl, & Hinton, 2013) is described in Equations (6) and (7):

$$u_{i+1} = \mu u_i - \alpha_L \nabla L(W_i + \mu u_i), \quad (6)$$

$$W_{i+1} = W_i + u_{i+1}, \quad (7)$$

where $L \in \{L^h, L^c\}$ is the loss function, μ is the momentum value, α_L is the learning rate, u_i is the momentum variable, ∇ is the rate of change in L , i is the iteration number and W_i denote the learnable weights. We employed randomly initialized weights for the scratch CNN and pretrained weights from the ImageNet dataset for the fine-tuned CNN (GoogleNet architecture). In addition to our modification, we remark that the original GoogleNet (in the Caffe framework) uses a simple online data augmentation that involves cropping (with a default crop size of 224×224 pixels), i.e. cutting out several patches from an input image at five positions (as five in a dice), and additionally flipping (horizontal reflection) to obtain more samples. During training of the CNN model, it automatically flips each cropped image to double the effective dataset size. The cropping means an act of extracting some portions from an input image. In our customized CNNs, we considered the original and two additional crop sizes: 125×125 and 250×250 pixels. The crop size of 250×250 implies the single actual size of the input image. Furthermore, we evaluated flip and non-flip conditions. All the input images to the CNN have image sizes of 250×250 pixels. For the ROT-DA images, each cell of the 4×4 grid contains a copy of the input image in a reduced size and the method fills up empty spaces with nearest neighbour pixels.

The derived three inception module CNN architecture is described in Table 2. This architecture involves the use of three inception modules that allow the concatenation of filters of different dimensions and sizes into a single new filter (Shin et al., 2016; Szegedy et al., 2015). In each inception module, there exist six convolution layers and one pooling layer. Moreover, there exist several rectifiers (ReLU) which are placed immediately after the convolutional and fully connected layers. Furthermore, there exist four pooling layers excluding those within the inception modules, two bottom convolutional layers and one top convolutional layer which comes after the average pooling layer. The authors in Lapin,

Table 2. Three inception module CNN architecture.

Layer type	Patch size/stride	Output size	Depth	Number of convolutional filters	Blob parameters
Conv 1	$7 \times 7/2$	$112 \times 112 \times 64$	1		16.06M
Max Pool 1	$3 \times 3/2$	$56 \times 56 \times 64$	0		4.01M
Conv 2	$3 \times 3/2$	$56 \times 56 \times 192$	2	0, 64, 192, 0, 0, 0	12.04M
Max Pool 2	$3 \times 3/2$	$28 \times 28 \times 192$	0		3.01M
Inception 3a		$28 \times 28 \times 256$	2	64, 96, 128, 16, 32, 32	4.01M
Inception 3b		$28 \times 28 \times 480$	2	128, 124, 192, 32, 96, 64	7.53M
Max Pool 3	$3 \times 3/2$	$14 \times 14 \times 480$	0		1.88M
Inception 4a		$14 \times 14 \times 512$	2	192, 96, 208, 16, 48, 64	2.01M
Average Pool 1		$4 \times 4 \times 512$	0		163.84K
Top Conv-1	$1 \times 1/1$	$4 \times 4 \times 128$	1		40.96K
FC 1 / 70% Drop L		$1 \times 1 \times 1024$	1 / 0		20.48K
FC 2		$1 \times 1 \times 2$	1		0.04K
CE / H Loss L		$1 \times 1 \times 2$	0		

Hein, and Schiele (2017) provide an analysis of loss functions for multi-class problems. We use a top-1 loss function which employs either the hinge loss or the cross-entropy loss (for the Softmax classifier). The L_1 -norm hinge loss L^h used in our study can be defined as

$$L^h(x_i) = \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K (\max(0, 1 - y_i^k z_k(x_i))), \quad (8)$$

where $y_i^k = \{1, -1\}$, $y_i^k = 1$, if x_i belongs to the target class of the k th class output unit, and $y_i^k = -1$ if x_i does not belong to the target class. The variable N denotes the total number of training images in a batch. K accounts for the number of class labels and $z_k = x^T w$ is the final activation of the output units. Here, $x \in \mathbb{R}^D$ denote the D -dimensional features of the previous hidden layer, and the learnable weights of the last layer are $w \in \mathbb{R}^{D \times K}$.

The cross-entropy loss L^c used in our study is defined as

$$L^c(x_i) = -\frac{1}{N} \sum_{i=1}^N y_i \left(\log \left(\frac{\exp(z_i(x_i))}{\sum_{k=1}^K \exp(z_k(x_i))} \right) \right), \quad (9)$$

where y_i denotes the target values $y_i \in \{0, 1\}$. The fraction within the log accounts for the softmax activation function (Okafor et al., 2016), which computes the probability distribution of the classes in a multi-class classification problem. Note that in this study, we investigate both binary and multi-classification problems.

The CNN under study consists of two fully connected (FC) layers: FC 1 with a corresponding ReLU computes the hidden unit activations, which is immediately followed by a regularization dropout of 0.7, and FC 2 contains the output neurons: 2, 12 and 6 for Aerial UAV, Croatia fish and Bird-600 datasets respectively. The working operations of the CNN are well explained in Szegedy et al. (2015).

3.2. Classical features combined with supervised learning algorithms

In this section, we describe the three feature extraction techniques which we use and combine with the k -nearest neighbour classifier and the SVM with a linear kernel or a radial basis function (RBF) kernel trained on the Aerial UAV dataset. In our preliminary experiments, we compared the classical approaches to the CNN techniques on the Aerial UAV dataset variants alone (without colour constancy). Note that for the classical techniques, we considered two image resolution sizes: 100×100 and 250×250 pixels. We remark that the classical methods performed worse compared to the CNN techniques. Hence, we only considered the CNN approach on the other two datasets. The classical methods are described as follows.

3.2.1. Colour histogram

The colour histogram (Colour Hist) is a feature extraction technique that analyses the pixel colour values within an image. For this, the pixel colour values of an image which exist as RGB (Red, Green and Blue) are first transformed to HSV (Hue, Saturation and Value). After that, the value of each pixel in a channel is put in a histogram consisting of different bins. In the experiments, only the saturation channel with a bin size of 32 is used, because it obtained the best performance in preliminary experiments. The

resulting feature vector containing 32 values is given to the supervised learning algorithms.

3.2.2. Histogram of oriented gradients

The histogram of oriented gradients (HOG) (Dalal & Triggs, 2005) features descriptor analyses patches (local regions) from an image. Then histograms are constructed based on the occurrences of orientation gradients within the patches. The HOG descriptor can process greyscale or colour image information. For the UAV dataset, we only considered the grey option. The procedure for constructing the HOG is as follows: convert the colour images of the aerial imagery into greyscale, then compute the gradients with two gradient kernels to compute the gradient values for each pixel from the greyscale image. The gradients for each pixel within a small block (cell) are put in bins (Junior, Delgado, Gonçalves, & Nunes, 2009; Karaaba, Surinta, Schomaker, & Wiering, 2015; Surinta, Karaaba, Schomaker, & Wiering, 2015; Takahashi, Takahashi, Cui, & Hashimoto, 2014), where each bin defines a specific orientation range. The following parameters were used, because they worked best in preliminary experiments: a grid of 2×2 blocks is used, where each block is split into 2×2 cells. The number of orientation bins is set to 4. This results in a feature dimension size of 64. This feature vector is fed as input to the supervised learning algorithms.

3.2.3. The combination of HOG and Colour Hist

In this technique, the features from both the HOG and Colour Hist are combined to form the HOG–Colour Hist feature descriptor. The features from both the HOG and Colour Hist are first computed separately. The optimal parameters used for HOG in the combined feature are different from the HOG descriptor alone, because they gave slightly better results in the preliminary experiments. The HOG parameters used in this technique use 32×32 pixels per cell, for which we used 9 cells in total from 100×100 pixel images with a single block. The number of orientation bins is set to 4 and the final feature dimensionality is 36. We used the hue channel from the colour histogram with 32 bins. These features are normalized and concatenated to obtain the final feature vector with 68 elements.

Several experiments were conducted to determine the best choice of parameters for the used classifiers with the different classical feature descriptors. For the k parameter in k -nearest neighbour (KNN), we tried $k = \{1, 2, 3, 4, 5, 10\}$. The C parameter of the linear SVM is set to $C = 2^{q-1}$, with the explored values $q \in \{1, 2, \dots, 19\}$. For the SVM with the RBF kernel, we tried $C = \{1, 2, 3, 5\}$ with $\gamma = 10^{p-1}$, where $p \in \{1, 2, \dots, 4\}$. The optimal parameters used for each of the classifiers are reported in Table 3. All the algorithms used for the classical techniques were developed in Python.

4. Experimental setup and results

This section entails the description of the experimental setup and shows and discusses the results on the used datasets.

4.1. CNN experimental setup

In this section, we explain the experimental setups used for each of the datasets.

Table 3. Best found parameters used for the various classifiers with the classical feature descriptors for the Aerial UAV dataset.

Classical techniques	RBF-SVM	Linear SVM	K-NN
HOG	$C=3, \gamma = 1000$	$C=8$	$K=1$
Colour Hist	$C=1, \gamma = 100$	$C=8192$	$K=3$
HOG–Colour Hist	$C=1, \gamma = 100$	$C=256$	$K=3$

4.1.1. CNN experimental setup for the Aerial UAV dataset

The enumeration below briefly describes the CNN setups for the experiments without and with colour constancy DA variants of this dataset.

- (1) CNN setup on the non-colour constancy DA variants of the Aerial UAV dataset: All experiments were run on the Caffe deep learning framework on a Ge-Force GTX 960 GPU model. The used experimental parameters are as follows: training display interval is set to 40, average loss is set to 40, learning rate is set to 0.001, learning policy is set to step, the step size is set to 4000 iterations, power is set to 0.5, gamma is set to 0.1, the momentum value is set to 0.9, weight decay is set to 0.0002 and maximum iteration is set to 10,000, which generates a snapshot model after every 500 iterations (which represent a snapshot). This resulted in 20 snapshots for the entire training process. The mentioned parameters were not altered during all the experiments for the different model configurations. The training images from the combination of any of the two subsets as reported in Table 1 are further split into the ratio 80% for training and 20% for validation. We employed a training batch size set to 20 and testing batch size set to 5 for all experiments, but with different test iterations. The altered parameters for the three subsets of the Aerial UAV dataset used with their corresponding splits are described in Table 4.

We first performed experiments with both the original and our derived CNN trained from scratch on the ORIG images. The preliminary results show that our proposed architecture requires less memory usage and a decrease in training computing time. This is summarized in Table 5. Additionally, our architecture obtains a similar level of performance compared to the original CNN.

- (2) CNN setup on the colour constancy DA variants of the Aerial UAV dataset: In this dataset, the effective sizes of the train-validation sets of the variants of colour constancy DA images in either original or rotation matrix form are increased to double or three times the original dataset size for the different subsets of this dataset. The new versions of the datasets result in a slight modification of the CNN training parameters: changes in the solver test iterations (validation/train) for the respective

Table 4. CNN parameters and dataset split information.

Parameters	Subset 1	Subset 2	Subset 3
Test images	262 (~ 7%)	2569 (~ 65%)	1150 (~ 29%)
Training images	2975 (~ 74%)	1129 (~ 28%)	2264 (~ 57%)
Validation images	744 (~ 19%)	283 (~ 7%)	567 (~ 14%)
Total images	3981 (100%)	3981 (100%)	3981 (100%)
Solver test iteration (val/train)	148	56	113
Test iterations for evaluation	52	514	230

Table 5. Preliminary experiment using original and our proposed CNN on the three cross-set splits of the Aerial UAV dataset.

Evaluation/methods	Derived CNN, NAGD	Original CNN, NAGD
Time (min)	$25.1 \leq t \leq 26.8$	$63.2 \leq t \leq 69.1$
Memory usage (MB)	752	1079
Average validation (%)	99.94	99.94
Average test (%)	97.87	97.71
Time improvement (%)	61.3 (decrease)	–

datasets are detailed in Table 6. The table also shows the dataset distribution. Moreover, we employed similar experimental settings as explained before. We remark that the test iterations for the three test sets that exist in either ORIG or ROT-DA alone were kept constant with the aim to examine the effectiveness of the new CNN models. Please note that we separated the rotation matrix and original versions of the test sets before applying colour constancy only on the train validation sets.

4.1.2. CNN experimental setup for Croatia fish dataset

In this dataset, we investigated the ORIG and ROT-DA datasets alone, and colour constancy data augmentation of ORIG and ROT-DA separately. Moreover, we also studied the impact of grid resolution on the ROT-DA; this means we used 4×4 and 2×2 ROT-DA images in our experiments separately. Similar CNN experimental settings as described in Section 4.1.1 were used. The additional modifications to the proposed CNN include the batch size for training, validation and testing is set to 12/8/1 respectively. The training of each of the CNN models uses maximum iterations of 7200, which generates a snapshot at each interval of 720 iterations, the step-size is set to 3600. This results in a decrease in the learning rate to 1/10th times the base learning rate of 0.001. For the ORIG and ROT-DA alone, we set the test interval to 240 while for the colour constancy DA versions (ORIG/ROT-DA) it is set to 720. The dataset variants were shuffled based on fivefold cross-validation with five different test sets ensuring no overlap exists in the train validation sets. Please note that we separated the rotation matrix and original versions of the test sets before applying colour constancy only on the train validation sets. The dataset distributions are detailed in Table 6.

Table 6. Dataset split information.

Dataset	Dataset variants	Sub	Train	Val	Test	STI
UAV	ROT+MSR-ROT-DA	Sub 1	5950	1488	262	297
	ROT+MSRCR-ROT-DA	Sub 2	2259	565	2569	113
	ORIG+MSR-ORIG-DA	Sub 3	4529	1133	1150	226
	ORIG+MSRCR-ORIG-DA					
	ROT+MSRCR-ROT+MSR-ROT-DA	Sub 1	8925	2232	262	446
	ORIG+MSRCR-ORIG+MSR-ORIG-DA	Sub 2	3388	848	2569	169
Bird		Sub 3	6793	1700	1150	340
	ORIG, ROT-DA	5 folds	270	30	300	30
	ROT+MSRCR-ROT+MSR-ROT-DA	5 folds	810	90	300	90
	ORIG+MSRCR-ORIG+MSR-ORIG-DA	5 folds	810	90	300	90
Fish	ORIG, ROT-DA	5 folds	240	160	394	20
	ROT+MSRCR-ROT+MSR-ROT-DA	5 folds	720	480	394	60
	ORIG+MSRCR-ORIG+MSR-ORIG-DA	5 folds	720	480	394	60

For the Aerial UAV dataset, the first four DA methods construct a dataset two times larger than the original dataset for all subfolds. STI means solver test iterations.

4.1.3. CNN experimental setup for Bird-600 dataset

In this dataset, we investigated the ORIG and ROT-DA alone, and colour constancy data augmentation of ORIG and ROT-DA separately. Our preliminary experiments suggest that the 2×2 ROT-DA yields better performances as compared to the larger 4×4 grid. This informed our choice of this grid, so we will use smaller grids for this dataset. A similar CNN experimental setup as described in Section 4.1.1 is used. The additional modification to the proposed CNN includes the batch size for training, validation and testing is set to 9/1/1 respectively. The training of each of the CNN models uses maximum iterations of 8100, which creates a snapshot at each interval of 810 iterations, the step-size is set to 4000. We used a base learning rate of 0.001. For the ORIG and ROT-DA alone, we set the test interval to 270 while for the colour constancy DA versions (ORIG/ROT-DA) it is set to 810. Similarly, the various dataset variants were shuffled based on fivefold cross-validation with five different test sets ensuring no overlap exists in the train validation set. Please note that we separated the rotation matrix and original versions of the test sets before applying colour constancy only on the train validation sets. The dataset distributions are detailed in Table 6.

4.2. Evaluation of the CNN architecture on the datasets

In this section, we discuss the classification performances on the used datasets.

4.2.1. Results on the Aerial UAV dataset

To compute the average results of the different subsets of this dataset, we compute the weighted average accuracy, which is computed by summing over the relative testing dataset sizes multiplied with the average accuracies on the testing datasets. The weighted mean can be computed using the expression: $T_m = (\sum_{s=1}^S W_s T_s) / (\sum_{s=1}^S W_s)$, where T_m denotes the weighted mean test accuracies, W_s denote the weights, which represent the number of individual images per test subset $W_s = \{262, 2569, 1150\}$, and T_s are the test accuracies for the various subsets, with $S=3$.

- (1) Evaluation of the CNN on Aerial UAV dataset variants (without colour constancy): In our preliminary studies, we carried out experiments on the data augmentation (ROT-DA) version of our dataset to determine the optimal crop size. We used models generated from the train validation experiments for evaluating our test sets. We initially employed the scratch CNN with the cross-entropy classification loss, which is combined with or without flipping and with different crop sizes: 125×125 , 224×224 and 250×250 . The results of these experiments are shown in Figure 9(a) and suggest that the optimal method uses a crop size of 224×224 pixels with flipping. This yields an accuracy of 98.18% that occurred at the 5th snapshot. We observed in general that there exist marginal differences between the various settings.

Based on this outcome, we used the best crop size with flip settings to carry out the experiments using the scratch and fine-tuned versions of the CNN. For this, we used both the ROT-DA and ORIG images. The validation results from Figure 9(b) show that the scratch and the fine-tuned CNN applied on the two kinds of images converge to a near maximum level of performance. The reason for this lies in the fact that most of the validation images contain similar objects as in the training set. The validation results at

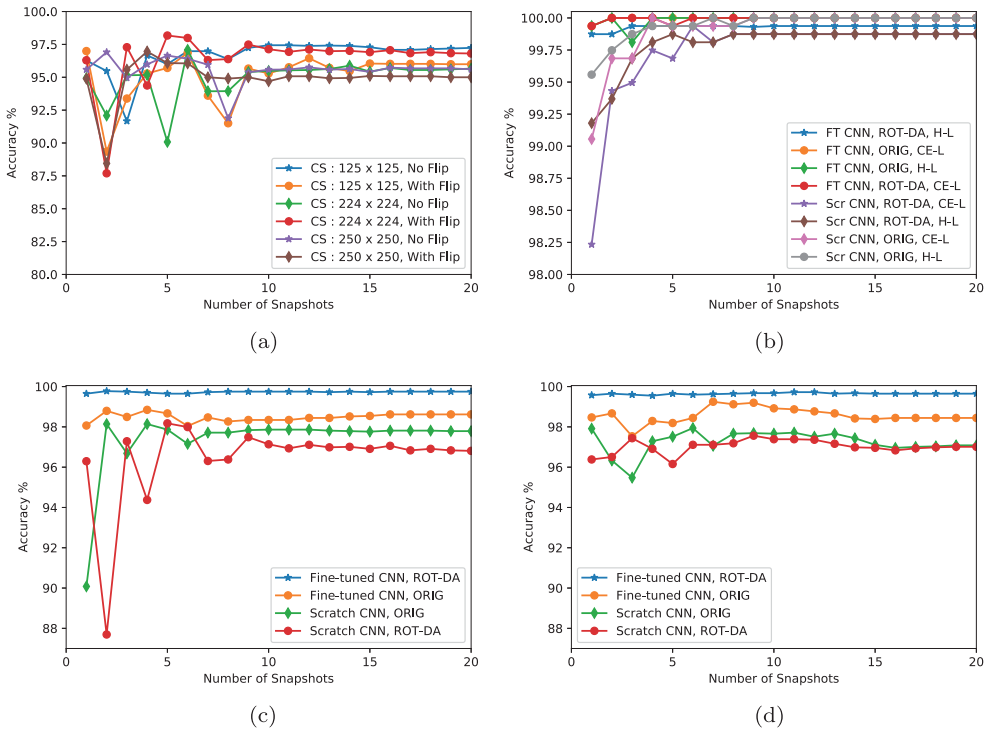


Figure 9. Weighted mean classification accuracy on the Aerial UAV dataset while training for 10,000 iterations (20 snapshots). (a) Preliminary test performance using scratch CNN with cross-entropy loss (softmax classifier) applied on ROT-DA alone using different crop sizes (CS), with and without flips. (b) Validation set evaluation of the CNN with cross-entropy loss (CE-L) and hinge loss (H-L) using a crop size of 224×224 and flip. The ROT-DA means the augmented dataset and ORIG means the originally up-scaled images. The FT means fine-tuned and Scr means Scratch. (c) Test evaluation of the CNN with CE loss using a crop size of 224×224 and flip. (d) Test evaluation of the CNN with L1-Norm hinge loss using a crop size of 224×224 and flip.

the 5th snapshot are reported in Table 7. From the table, we can see that the use of the original dataset leads to more overfitting. The results of the different CNNs with the cross-entropy loss function are shown in Figure 9(c). From this figure, we can observe that the best obtained test accuracy is obtained by the fine-tuned CNN applied on the ROT-DA images in the 2nd snapshot. We further investigated the CNN with the L_1 hinge loss, using the earlier mentioned CNN settings (scratch and

Table 7. Weighted mean of the test and validation classification accuracies of the CNN applied on the Aerial UAV dataset after the 5th snapshot.

Evaluation	Method	Cross-entropy loss	Hinge loss
Test	Fine-tuned CNN, ROT-DA	99.65	99.65
	Fine-tuned CNN, ORIG	98.67	98.19
	Scratch CNN, ROT-DA	98.18	96.16
	Scratch CNN, ORIG	97.87	97.51
Validation	Fine-tuned CNN, ROT-DA	99.94	99.94
	Fine-tuned CNN, ORIG	100.00	100.00
	Scratch CNN, ROT-DA	99.68	99.81
	Scratch CNN, ORIG	99.94	99.94

fine-tuned versions) applied on the two sets of images (ROT-DA and ORIG). The results obtained are shown in [Figure 9\(d\)](#).

Based on the performances recorded during this preliminary investigation, we only compared results obtained at the 5th snapshot as reported in [Table 7](#). The results show that the fine-tuned CNN trained on the data-augmented images yields higher test classification accuracies when compared to the fine-tuned CNN trained on the ORIG images of the dataset. We compared the different approaches using the binomial distribution of correctly classifying test images. The results show that the fine-tuned CNN trained on the data-augmented images yields significantly higher classification accuracies ($P < 0.01$) when compared to the fine-tuned CNN trained on the ORIG images of the dataset. Overall, the fine-tuned CNNs obtain the best results and combined with the data-augmented images, the results are very good (99.65%). Finally, the results show that overall the use of the cross-entropy loss function leads to better results than the use of the hinge loss function.

- (2) Evaluation of classical descriptors on the Aerial UAV dataset (without colour constancy): The weighted mean test accuracies of the classical techniques on the Aerial UAV dataset are reported in [Table 8](#). We observe that the RBF-SVM outperforms the other two classifiers (K-NN and linear SVM) when combined with each of the feature descriptors. Another observation is that the classifiers with the Colour Hist or HOG–Colour Hist features yield better performances than using the HOG descriptor alone.

This shows the importance of using colour information for this classification problem. Still, the results are significantly worse than the results using the CNN

Table 8. Summary of the weighted mean test performances for all CNNs and the classical methods on the Aerial UAV dataset.

Methods	Sub 1	Sub 2	Sub 3	Weighted mean
Fine-tuned CNN, ROT-DA, cross-entropy loss	100.00	99.73	99.39	99.65
Fine-tuned CNN, ROT-DA, hinge loss	99.62	99.77	99.39	99.65
Fine-tuned CNN, ORIG, cross-entropy loss	99.62	98.29	99.30	98.67
Fine-tuned CNN, ORIG, hinge loss	99.62	97.55	99.30	98.19
Scratch CNN, ROT-DA, cross-entropy loss	98.23	98.72	96.96	98.18
Scratch CNN, ROT-DA, hinge loss	98.08	96.19	95.65	96.16
Scratch CNN, ORIG, cross-entropy loss	98.85	99.34	94.35	97.87
Scratch CNN, ORIG, hinge loss	97.69	98.83	94.52	97.51
RBF-SVM-HOG, ORIG-100×100	96.56	86.99	95.30	90.02
RBF-SVM-Colour Hist, ORIG-100×100	96.56	96.07	96.87	96.33
RBF-SVM-HOG-Colour Hist, ORIG-100×100	96.56	96.11	96.69	96.31
Linear SVM-HOG, ORIG-100×100	85.88	81.51	95.65	85.88
Linear SVM-Colour Hist, ORIG-100×100	96.95	93.77	95.83	94.57
Linear SVM-HOG-Colour Hist, ORIG-100×100	95.80	94.08	93.74	94.09
KNN-HOG, ORIG-100×100	88.17	84.35	96.78	88.19
KNN-Colour Hist, ORIG-100×100	96.56	96.50	94.86	96.03
KNN-HOG-Colour Hist, ORIG-100×100	96.95	96.46	94.78	96.01
RBF-SVM-HOG, ORIG-250×250	85.88	81.51	95.65	85.88
RBF-SVM-Colour Hist, ORIG-250×250	96.57	95.37	96.52	95.78
RBF-SVM-HOG-Colour Hist, ORIG-250×250	85.88	81.51	95.65	85.88
RBF-SVM-HOG, ROT-DA-250×250	85.88	81.51	95.65	85.88
RBF-SVM-Colour Hist, ROT-DA-250×250	96.18	95.25	96.70	95.73
RBF-SVM-HOG-Colour Hist, ROT-DA-250×250	95.04	93.97	96.08	94.65
RBF-SVM-HOG-ROT-DA-NR-250×250	94.66	81.51	86.61	83.84
RBF-SVM-Colour Hist-ROT-DA-NR-250×250	96.56	95.13	96.43	95.60
RBF-SVM-HOG-Colour Hist-ROT-DA-NR-250×250	95.04	91.98	96.43	93.47

methods. Table 8 also shows the results of using the RBF-SVM with different datasets and different feature descriptors using larger images (250×250 pixels). The results show that here data augmentation does not lead to significantly better results. This can be explained by the fact that the best feature descriptor, the colour histogram, is not affected by this DA method. Finally, we note that the ORIG image with the smaller 100×100 resolution works better for the HOG feature descriptor and therefore also for HOG combined with the colour histogram. This can be explained by the fact that we optimized the HOG parameters using the smaller images.

Although the performances of the CNN techniques are much better, the classical techniques have a lower training computing time: $t \leq 1$ min. This is because of the low dimensionality of the extracted features and the low number of trainable parameters.

- (3) Results of the CNN on the Aerial UAV dataset variants (with colour constancy): The CNN training computing time on the colour constancy DA variants for the different subsets is $t \leq 46$ min. We used the same approach of computing the weighted mean of the accuracies for the three subsets as reported before. The subfigures in Figure 10 show the learning curves for both training and testing on the colour constancy DA variants of ORIG and ROT images respectively. From Figure 10(a,b), we

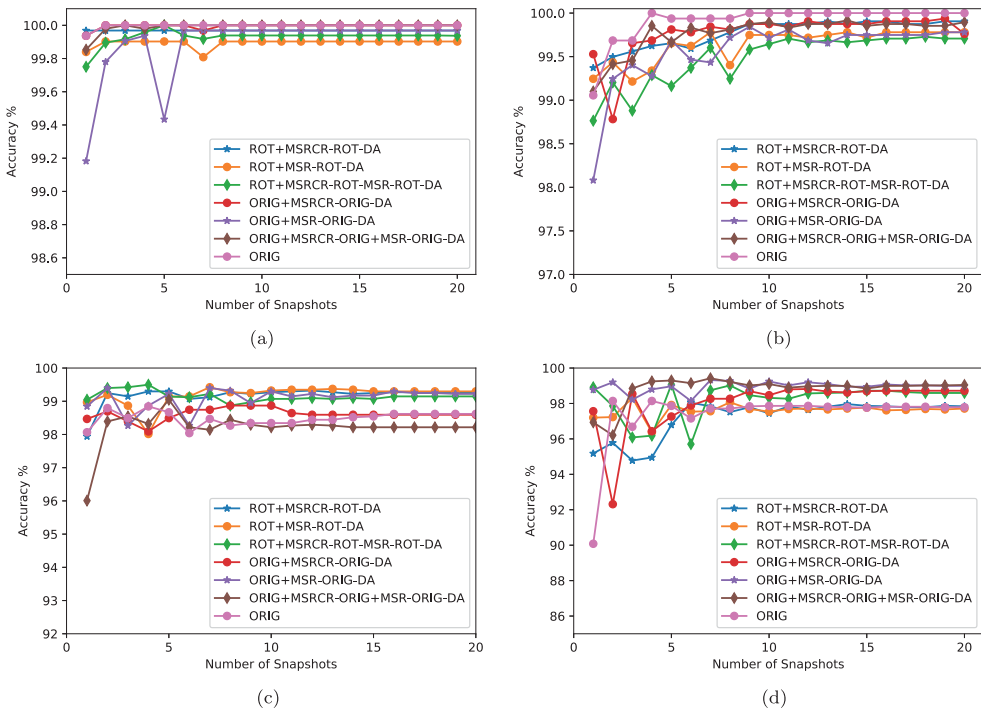


Figure 10. Weighted mean classification accuracy on the Aerial UAV dataset (different colour constancy DA approaches) while training for 10K iterations (20 snapshots) using CNN with cross-entropy loss function. Please note that not all graphs are visible due to overlap. (a) Validation evaluation of the fine-tuned CNN, (b) validation evaluation of the scratch CNN, (c) test evaluation of the fine-tuned CNN and (d) test evaluation of the scratch CNN.

observe that CNN validation accuracies of the colour constancy DA methods yield very similar performances for both fine-tuned and scratch experiments.

From Figure 10(c), we observe that the use of fine-tuned CNN on the ROT-MSRCR-ROT+MSR-ROT-DA attained a peak accuracy of $\sim 99.5\%$ at the 4th snapshot while that of fine-tuned CNN on the ORIG+MSRCR-ORIG+MSR-ORIG obtained $\sim 99.06\%$ at the 5th snapshot. In both approaches, the performances reduce for longer iterations; this suggests that early stopping will be most appropriate for these methods. The validation performance in Figure 10(a) shows that most of the techniques examined were stable after the 7th snapshot (3.5K iterations). Hence we choose this iteration point as the basis of our comparison. A summary of the validation and the test accuracies is reported in Table 9. Overall, the fine-tuned CNN applied on ROT+MSR-ROT-DA yields a very good performance for almost all iterative points of evaluation.

In this dataset, using fine-tuned CNN on colour constancy data augmentation with ROT images yields a higher accuracy than with the fine-tuned CNN using either colour constancy data augmentation with ORIG images or ORIG images alone. However, all fine-tuned CNN results obtained using colour constancy DA images do not surpass results obtained from fine-tuned CNN on ROT-DA images alone. This is possibly due to fact that the test sets are only using ROT-DA images. Overall the proposed rotation matrix algorithm leads to higher accuracies on this dataset with or without the colour constancy algorithm.

In contrast to this observation, in the scratch experiments, the results obtained from training scratch CNNs on colour constancy data augmentation with ORIG images outperform CNN results obtained on ROT-DA and ORIG images alone. Thus it seems that adding more images to train the scratch CNNs plays the most important role. Based on this observation, we will use the best scratch technique (ORIG-MSRCR-ORIG+MSR-ORIG-DA) and its rotation matrix version on the next two datasets. It is surprising that the scratch CNN performs better than the fine-tuned CNN on the ORIG+MSRCR-ORIG+MSR-ORIG-DA dataset.

Table 9. Weighted mean of the validation and test classification accuracies of the CNN applied on different versions of the Aerial UAV dataset.

Training CNN	Dataset variants	Validation	Test
Fine-tuned CNN	ROT-DA (Okafor et al., 2017)	99.94	99.65
	ROT+MSR-ROT-DA	99.81	99.42
	ORIG+MSR-ORIG-DA	99.97	99.40
	ROT+MSRCR-ROT+MSR-ROT-DA	99.91	99.22
	ROT+MSRCR-ROT-DA	99.97	99.12
	ORIG+MSRCR-ORIG-DA	99.97	98.74
	ORIG (Okafor et al., 2017)	100.00	98.67
	ORIG+MSRCR+ORIG+MSR-ORIG-DA	100.00	98.14
	ORIG+MSRCR+ORIG+MSR-ORIG-DA	99.73	99.41
	ORIG+MSR-ORIG-DA	99.43	99.32
Scratch CNN	ROT+MSRCR-ROT+MSR-ROT-DA	99.60	98.74
	ORIG+MSRCR-ORIG-DA	99.84	98.27
	ROT-DA (Okafor et al., 2017)	99.68	98.18
	ORIG (Okafor et al., 2017)	100.00	97.87
	ROT+MSRCR-ROT-DA	99.69	97.84
	ROT+MSR-ROT-DA	99.77	97.56

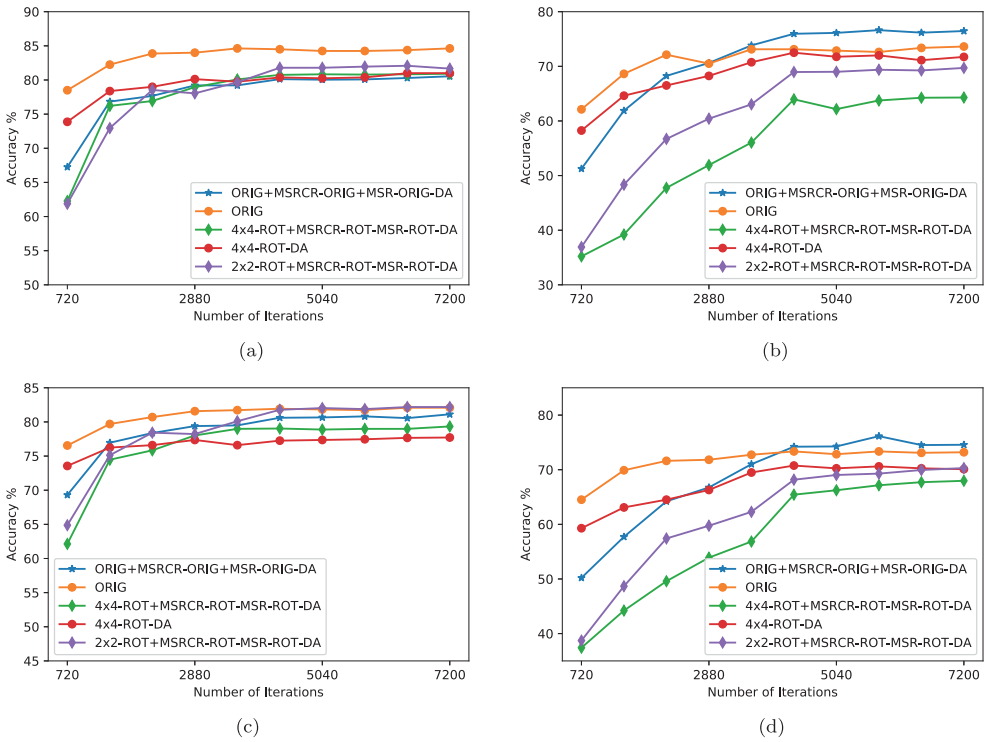


Figure 11. Fivefold cross-validation mean classification accuracy on the Croatia fish dataset while training for 7200 iterations using CNNs with the cross-entropy loss function: (a) validation evaluation of the fine-tuned CNN, (b) validation evaluation of the scratch CNN, (c) test evaluation of the fine-tuned CNN and (d) test evaluation of the scratch CNN.

This may be caused by some overfitting problem, which we observed in the test accuracy of subset 2.

4.2.2. Results on Croatia fish dataset

We trained the CNNs using fivefold cross-validation data splits. The training time of the CNN models for each of the methods is $t \leq 16$ min. The models generated from the CNNs using colour constancy DA variants with (ROT or ORIG) or (ROT or ORIG alone) were used to compute the accuracy on the test sets that contain either ORIG or ROT-DA images without colour constancy. The learning curves for train validation and testing phases while training for 7200 iterations are shown in Figure 11.

The mean accuracies for test and validation sets for the different approaches after that number of iterations are reported in Table 10. We report that there is no significant difference between the test and validation performances for most methods. This indicates that the test and validation performances are consistent.

From Table 10, we observe that the fine-tuned CNN on ORIG alone, the colour constancy data augmentation on ORIG and the 2×2 -ROT version of the dataset all yield high accuracies. There is no significant difference in accuracies between these three methods. The best method is the fine-tuned CNN on the 2×2 -ROT+MSRCR-ROT+MSR-

Table 10. Fivefold cross-validation and test classification accuracies and standard deviations of the CNN applied on different versions of the Croatia fish dataset.

Training CNN	Dataset variants	Validation	Test
Fine-tuned CNN	2×2 -ROT+MSRCR-ROT+MSR-ROT-DA	81.67 ± 2.65	82.18 ± 3.44
	ORIG	84.63 ± 2.78	82.08 ± 4.21
	ORIG+MSRCR+ORIG+MSR-ORIG-DA	80.54 ± 2.81	81.12 ± 3.16
	4×4 -ROT+MSRCR-ROT+MSR-ROT-DA	80.92 ± 2.82	79.34 ± 1.66
	4×4 -ROT-DA	81.00 ± 1.66	77.72 ± 1.72
Scratch CNN	ORIG+MSRCR+ORIG+MSR-ORIG-DA	76.46 ± 2.40	74.56 ± 4.10
	ORIG	73.64 ± 1.99	73.19 ± 3.12
	2×2 -ROT+MSRCR-ROT+MSR-ROT-DA	69.71 ± 3.43	70.30 ± 4.20
	4×4 -ROT-DA	71.73 ± 3.08	70.10 ± 3.74
	4×4 -ROT+MSRCR-ROT+MSR-ROT-DA	64.29 ± 1.30	67.97 ± 4.88

ROT-DA variant of this dataset. When we compare the results of the fine-tuned CNN applied on 2×2 -ROT+MSRCR-ROT+MSR-ROT-DA to 4×4 -ROT-DA, there exists a significant difference ($P < 0.05$). This indicates that the use of colour constancy data augmentation with ROT images and the right choice of grid resolution are important for this dataset. We also note that the fine-tuned CNN significantly outperforms the scratch CNN on this dataset.

For the scratch experiments, training the CNN using ORIG+MSRCR-ORIG+MSR-ORIG-DA yields the highest accuracy. This best scratch CNN approach significantly outperforms the 4×4 ROT+MSRCR-ROT+MSR-ROT ($P < 0.05$). Overall, the choice of colour constancy data augmentation with 2×2 ROT images works better in our experiment than the use of colour constancy data augmentation with 4×4 ROT images.

4.2.3. Results on bird dataset

We trained the CNNs using fivefold cross-validation data splits. The training time of the CNN models for each of the methods is $t \leq 13$ min. The models generated from the CNNs using colour constancy DA variants with (ROT or ORIG) or (ROT or ORIG alone) of this dataset were used to compute the accuracies on the test sets that only contain either ORIG or ROT-DA images (without colour constancy images). The learning curves for train validation and testing phases, while training for 8100 iterations are shown in Figure 12. The mean accuracies for test and validation sets after that number of iterations are reported in Table 11. From this table, we report that there is no significant difference between the test and validation performances for each of the examined methods, this shows again that the test and validation performances are consistent to each other. From the subfigures in Figure 12, we observe that the fine-tuned CNNs outperform the scratch CNN methods on the different dataset variants.

The best techniques are the fine-tuned CNN on either 1×1 -ROT+MSRCR-ROT+MSR-ROT-DA or ORIG+MSRCR+ORIG+MSR-ORIG-DA. These results indicate the importance of colour constancy on the ROT or ORIG images. This success can be attributed to training CNN weights with enhanced colour information and with more images. To obtain this better performance, it was important to choose smaller rotational bounds $[-15^\circ, 15^\circ]$ as used in the 1×1 -ROT+MSRCR-ROT+MSR-ROT-DA rather than the original rotational bounds $[1^\circ, 180^\circ]$. Such higher angular bounds may not be suitable for images that have an upright representation of objects.

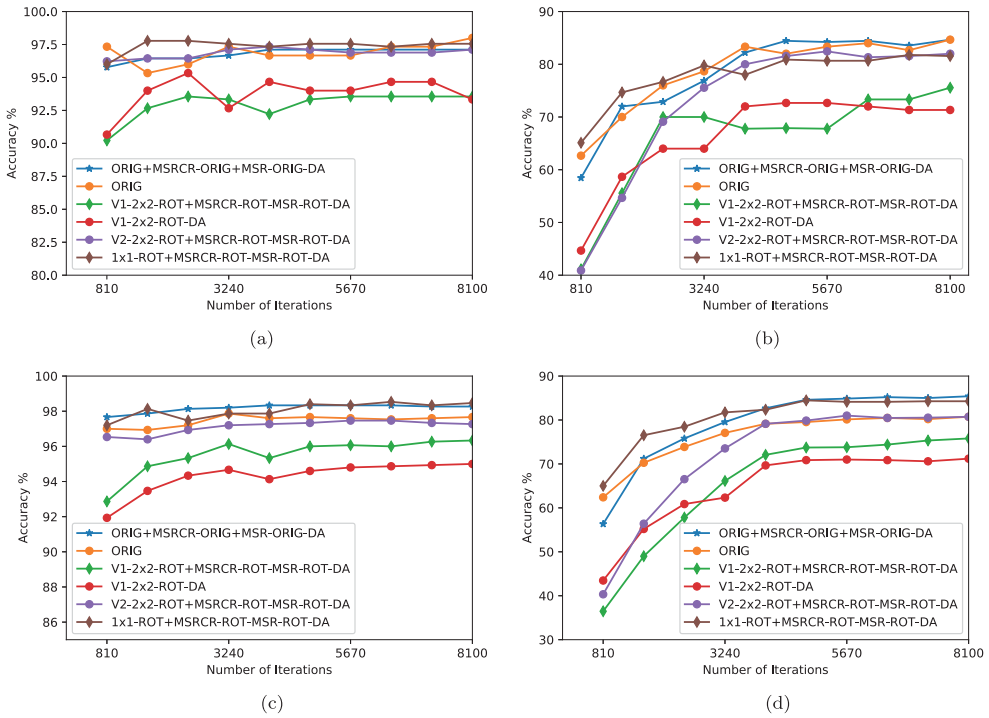


Figure 12. Fivefold cross-validation mean classification accuracy on the Bird-600 dataset while training for 8100 iterations using CNNs with the cross-entropy loss function: (a) validation evaluation of the fine-tuned CNN, (b) validation evaluation of the scratch CNN, (c) test evaluation of the fine-tuned CNN and (d) test evaluation of the scratch CNN.

Furthermore, we compared the results obtained with the fine-tuned CNNs on the different variants of this dataset to the baseline result from Lazebnik et al. (2005) which obtained 92.33% using a probabilistic part-based method (maximum entropy framework). Our best approach significantly outperformed the baseline with a margin of 6.14% using the fine-tuned CNN on 1×1 -ROT+MSRCR-ROT+MSR-ROT-DA. However, we remark that the obtained scratch CNN results on this dataset performed worse than the baseline method.

Table 11. Fivefold cross-validation and test classification accuracies and standard deviations of the CNN applied on different versions of the Bird-600 dataset.

Training CNN	Dataset variants	Validation	Test
Fine-tuned CNN	1×1 -ROT+MSRCR-ROT+MSR-ROT-DA	97.56 ± 2.47	98.47 ± 0.34
	ORIG+MSRCR+ORIG+MSR-ORIG-DA	97.11 ± 2.59	98.26 ± 0.25
	ORIG	98.00 ± 2.67	97.67 ± 0.56
	$V2 - 2 \times 2$ -ROT+MSRCR-ROT+MSR-ROT-DA	97.11 ± 2.29	97.27 ± 0.93
	$V1 - 2 \times 2$ -ROT+MSRCR-ROT+MSR-ROT-DA	93.55 ± 2.76	96.33 ± 0.79
	$V1 - 2 \times 2$ -ROT-DA	93.33 ± 2.98	95.00 ± 0.73
Scratch CNN	ORIG+MSRCR+ORIG+MSR-ORIG-DA	84.67 ± 5.23	85.40 ± 1.73
	1×1 -ROT+MSRCR-ROT+MSR-ROT-DA	81.56 ± 4.01	84.27 ± 1.58
	ORIG	84.67 ± 5.81	80.73 ± 2.73
	$V2 - 2 \times 2$ -ROT+MSRCR-ROT+MSR-ROT-DA	82.00 ± 5.94	80.73 ± 1.51
	$V1 - 2 \times 2$ -ROT+MSRCR-ROT+MSR-ROT-DA	77.11 ± 5.05	75.80 ± 1.15
	$V1 - 2 \times 2$ -ROT-DA	71.33 ± 7.48	71.20 ± 3.03

5. Conclusion

In deep learning, data augmentation can play an important role if a dataset does not contain many training images. In this paper, we developed a novel DA method that transforms an image into a new image containing multiple random transformations of the image. We combined this method with the use of colour constancy algorithms that add several transformed images to the training datasets. We created different combinations of methods: using ORIG or ROT images combined with colour constancy transformed images or not. These combinations were compared on three different animal datasets: Aerial UAV containing cows or not, a dataset with bird images and a dataset with fish images. Overall we considered two broad forms of data augmentation based on their increase (colour constancy data augmentation with ORIG or ROT-DA) or no increase (ROT-DA alone) in the amount of training images.

The results show that for the Aerial UAV dataset, the augmented ROT images are very useful. The Aerial UAV dataset consists of pictures taken from the sky, and therefore it is important to cope with 2D rotations to obtain the highest accuracies. It should be noted that this DA algorithm is useful for the CNNs, because although CNNs are more or less translational invariant, they are not rotational invariant. For the fish and birds dataset, the proposed rotation matrix DA method does not lead to better results than using the ORIG images. For these datasets, the images show objects which are often in an upright position, and therefore there is less need to battle rotational variances.

The colour constancy data augmentation helps in overall to get better accuracies, but the differences are not very large compared to using the ORIG images. Only on the bird dataset, the colour constancy data augmentation plays a very important role when training the CNN from scratch. The variation in colours is quite large for this dataset, and therefore adding additional images with different illumination levels is helpful. On this dataset, colour constancy data augmentation also improves the results of the fine-tuned CNN.

The results have also shown that the fine-tuned CNNs significantly outperform the CNNs trained from scratch on the Croatia fish and Bird-600 datasets. Furthermore, the fine-tuned CNNs obtain very high accuracies on the Aerial UAV and Bird-600 datasets.

Future works can explore the use of deep neural network architectures to artificially transform colours in images. This could be done with a novel way of data augmentation or by adding initial layers that immediately transform the colour pixels. It will also be interesting to create a deep neural network that can create the best ROT images, possibly trained using an adversarial learning framework.

Acknowledgments

This work acknowledges the Department of Artificial Intelligence and Cognitive Engineering and the Center for Information Technology at the University of Groningen for their support on this research.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Emmanuel Okafor is a PhD candidate at the Institute of Artificial Intelligence and Cognitive Engineering, the University of Groningen, the Netherlands. He earned a Master of Science in Control System Engineering from Ahmadu Bello University, Zaria. He is a lecturer for the past 5 years till date at the Department of Electrical and Computer Engineering, Ahmadu Bello University, Zaria, Nigeria. He is actively interested in research within the field of deep learning, computer vision and control systems.

Lambert Schomaker is a professor in perceptual intelligence and machine learning, currently the scientific director at the Department of Artificial Intelligence and Cognitive Engineering, University of Groningen, the Netherlands. He is a renowned professor in the field of handwriting recognition and has great interest in the fields of perceptual learning, robotics, computer-vision, deep learning and optimization.

Marco A. Wiering is currently an assistant professor in the field of machine learning in the Department of Artificial Intelligence and Cognitive Engineering at the University of Groningen, the Netherlands. He has successfully supervised or is supervising 10 PhD students and around 100 Master students. Dr. Wiering has co-authored more than 160 conference or journal papers in different fields such as computer vision, reinforcement learning, robotics, deep learning and optimization.

ORCID

Lambert Schomaker  <http://orcid.org/0000-0003-2351-930X>

References

- Caicedo, J. C., & Lazebnik, S. (2015). Active object localization with deep reinforcement learning. *Proceedings of the IEEE international conference on computer vision, Washington, DC* (pp. 2488–2496).
- Charalambous, C. C., & Bharath, A. A. (2016). A data augmentation methodology for training machine/deep learning gait recognition algorithms. *arXiv preprint arXiv:1610.07570*.
- Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *IEEE computer society conference on computer vision and pattern recognition (CVPR), San Diego, CA* (Vol. 1, pp. 886–893).
- Fang, Y., Du, S., Abdoola, R., Djouani, K., & Richards, C. (2016). Motion based animal detection in aerial videos. *Procedia Computer Science*, 92, 13–17.
- Galdran, A., Alvarez-Gila, A., Meyer, M. I., Saratzaga, C. L., Araújo, T., Garrote, E., ... Campilho, A. (2017). Data-driven color augmentation techniques for deep skin image analysis. *arXiv preprint arXiv:1703.03702*.
- Ghazi, M. M., Yanikoglu, B., & Aptoula, E. (2017). Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing*, 235, 228–235.
- Gonzalez, L. F., Montes, G. A., Puig, E., Johnson, S., Mengersen, K., & Gaston, K. J. (2016). Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation. *Sensors*, 16(1), 97.
- Imamura, Y., Okamoto, S., & Lee, J. H. (2016). Human tracking by a multi-rotor drone using HOG features and linear SVM on images captured by a monocular camera. *Proceedings of the international multiconference of engineers and computer scientists, Hong Kong* (Vol. 1).
- Jaeger, J., Simon, M., Denzler, J., Wolff, V., Fricke-Neuderth, K., & Kruschel, C. (2015). Croatian fish dataset: Fine-grained classification of fish species in their natural habitat. Swansea: Bmvc.
- Jeon, S., Shin, J-W., Lee, Y-J., Kim, W-H., Kwon, Y., & Yang, H-Y. (2017). Empirical study of drone sound detection in real-life environment with deep neural networks. *arXiv preprint arXiv:1701.05779*.
- Jiang, B., Woodell, G. A., & Jobson, D. J. (2015). Novel multi-scale retinex with color restoration on graphics processing unit. *Journal of Real-Time Image Processing*, 10(2), 239–253.

- Jobson, D. J., Rahman, Z.-u., & Woodell, G. A. (1997). A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7), 965–976.
- Junior, O. L., Delgado, D., Gonçalves, V., & Nunes, U. (2009). Trainable classifier-fusion schemes: An application to pedestrian detection. *Intelligent Transportation Systems, St. Louis, MO* (Vol. 2).
- Karaaba, M., Surinta, O., Schomaker, L., & Wiering, M. A. (2015). Robust face recognition by computing distances from multiple histograms of oriented gradients. *2015 IEEE symposium series on computational intelligence, Cape Town* (pp. 203–209).
- Katsigiannis, P., Misopolinos, L., Liakopoulos, V., Alexandridis, T. K., & Zalidis, G. (2016). An autonomous multi-sensor UAV system for reduced-input precision agriculture applications. *24th Mediterranean conference on control and automation (MED)*, Athens (pp. 60–64).
- Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1), 1–11.
- Lapin, M., Hein, M., & Schiele, B. (2017). Analysis and optimization of loss functions for multiclass, top-k, and multilabel classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Lazebnik, S., Schmid, C., & Ponce, J. (2005). A maximum entropy framework for part-based texture and object recognition. *ICCV'05. Tenth IEEE international conference on computer vision, 2005, Beijing* (Vol. 1, pp. 832–838).
- Lin, T.-Y., Cui, Y., Belongie, S., & Hays, J. (2015, June). Learning deep representations for ground-to-aerial geolocalization. *The IEEE conference on computer vision and pattern recognition (CVPR)*, Boston, MA.
- López-Granados, F., Torres-Sánchez, J., Serrano-Pérez, A., de Castro, A. I., Mesas-Carrascosa, F.-J., & Peña, J.-M. (2016). Early season weed mapping in sunflower using UAV technology: Variability of herbicide treatment maps against weed thresholds. *Precision Agriculture*, 17(2), 183–199.
- Lou, Z., Gevers, T., Hu, N., & Lucassen, M. P. (2015). Color constancy by deep learning. Swansea: Bmvc (pp. 76.1).
- Lukas, V., Novák, J., Neudert, L., Svobodova, I., Rodriguez-Moreno, F., Edrees, M., ... Kren, J. (2016). The combination of UAV survey and landsat imagery for monitoring of crop vigor in precision agriculture. *ISPRS – International archives of the photogrammetry, remote sensing and spatial information sciences* (pp. 953–957).
- Moore, A., Allman, J., & Goodman, R. M. (1991). A real-time neural system for color constancy. *IEEE Transactions on Neural networks*, 2(2), 237–247.
- Okafor, E., Pawara, P., Karaaba, F., Surinta, O., Codreanu, V., Schomaker, L. ... Wiering, M. (2016). Comparative study between deep learning and bag of visual words for wild-animal recognition. *IEEE symposium series for computational intelligence (SSCI)*, Athens.
- Okafor, E., Smit, R., Schomaker, L., & Wiering, M. (2017). Operational data augmentation in classifying single aerial images of animals. *2017 IEEE international conference on innovations in intelligent systems and applications (INISTA)*, Gdynia, Poland (pp. 354–360).
- Pawara, P., Okafor, E., Schomaker, L., & Wiering, M. (2017). Data augmentation for plant classification. *International conference on advanced concepts for intelligent vision systems, Antwerp* (pp. 615–626).
- Petro, A. B., Sbert, C., & Morel, J.-M. (2014). Multiscale retinex. *Image Processing On Line*, 4, 71–88.
- Provenzi, E., Marini, D., De Carli, L., & Rizzi, A. (2005). Mathematical definition and analysis of the retinex algorithm. *Journal of the Optical Society of America A*, 22(12), 2613–2621.
- Rahman, Z.-u., Jobson, D. J., & Woodell, G. A. (1996). Multi-scale retinex for color image enhancement. *International conference on image processing, 1996. Proceedings, Lausanne* (Vol. 3, pp. 1003–1006).
- Rahman, Z.-u., Jobson, D. J., & Woodell, G. A. (2004). Retinex processing for automatic image enhancement. *Journal of Electronic Imaging*, 13(1), 100–110.
- Rogez, G., & Schmid, C. (2016). MoCap-guided data augmentation for 3D pose estimation in the wild. *Advances in neural information processing systems* (pp. 3108–3116). Barcelona: Curran Associates, Inc.
- Salehinejad, H., Valaee, S., Dowdell, T., & Barfett, J. (2018). Image augmentation using radial transform for training deep neural networks. *IEEE International Conference on Acoustics, Speech and Signal Processing, Calgary, Alberta, Canada*.

- Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ... Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5), 1285–1298.
- Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience*, 2016, 1–11.
- Surinta, O., Karaaba, M. F., Schomaker, L. R., & Wiering, M. A. (2015). Recognition of handwritten characters using local gradient feature descriptors. *Engineering Applications of Artificial Intelligence*, 45, 405–414.
- Sutskever, I., Martens, J., Dahl, G. E., & Hinton, G. E. (2013). On the importance of initialization and momentum in deep learning. *International Conference on Machine Learning ICML (3)*, 28, 1139–1147.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D. ... Rabinovich, A. (2015). Going deeper with convolutions. *The IEEE conference on computer vision and pattern recognition CVPR*, Boston, MA.
- Takahashi, K., Takahashi, S., Cui, Y., & Hashimoto, M. (2014). Remarks on computational facial expression recognition from HOG features using quaternion multi-layer neural network. *Engineering applications of neural networks* (pp. 15–24). Sofia: Springer.
- Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: A review. *Precision Agriculture*, 13(6), 693–712.